Doble Grado en II+ADE 04/06/ 2019

Ecometría Examen Final Curso 2018-2019

DNI/NIE : APELLIDOS: Nombre:

Cada pregunta se resuelve en la hoja de su enunciado: no se pueden responder preguntas distintas en la misma hoja. Las respuestas se deben escribir con tinta azul o negra. Las respuestas deben ser breves pero razonadas. Errores conceptuales importantes pueden afectar a la calificación global del examen.

Teoría(3 puntos). 1) Enunciar el teorema de Gauss-Markov en regresión simple y múltiple. ¿Cúal es su relevancia?

2) Demostrar el teorema de Gauss-Markov en regresión simple, a partir del teorema de Gauss-Markov en regresión múltiple.

3) ¿Qué hipótesis del modelo de Gauss-Markov no son necesarias para la demostración del teorema de Gauss-Markov?

4) ¿Existe algún test F que sea equivalente a un test t en regresión lineal?, si la respuesta es afirmativa, dar una justificación informal.

Aunque las respuestas deberían estar claras a partir de la teoría, damos las soluciones de 2), 3) y 4).

2) En la demostración del teorema se verifica que $(\boldsymbol{\gamma}) - \text{var}(\hat{\boldsymbol{\beta}}) = \sigma^2 BB'$. Para $k = 2$, $B = (\mathbf{b}_1\, \mathbf{b}_2)$ entonces

$$BB' = \begin{pmatrix} |\mathbf{b}_1|^2 & \mathbf{b}_1 \cdot \mathbf{b}_2 \\ \mathbf{b}_1 \cdot \mathbf{b}_2 & |\mathbf{b}_2|^2 \end{pmatrix},$$

ie, matriz de productos escalares. Por tanto, los elementos de la diagonal de $\sigma^2 BB'$ no pueden ser negativos, en particular, no lo será la diferencia de las varianzas de los dos estimadores correspondientes del coeficiente β_2, que es lo que afirma el teorema en regresión simple.

3) Para demostrar que $\hat{\boldsymbol{\beta}}$ es insesgado (necesario a su vez para obtener la expresión de su matriz de covarianzas) necesitamos (H1) (linealidad), (H2) ($\text{rang}(X) = k$), (H3) (perturbación con media cero). Para obtener la expresión de la matriz de covarianzas $\text{var}\hat{\boldsymbol{\beta}}$, necesitamos además (H4) (homocedasticidad) y (H5) (no correlación dos a dos en las perturbaciones). Por tanto, la única hiportesis que no ncesitamos es (H6), que las perturbaciones son normales: por esa razón a veces no se incluye en el modelo.

4) El test t_i de significación individual de una variable con H0: $\beta_i = 0$ es equivalente al test F para contrastar al suprimir la única variable la variable β_i,con la misma hipóteis $H0$; en particular los p-valores serán los mismos. La justificación del test se encuentra en la relación entre las distribuciones de las variables aleatorias correspondientes: $t^2_{n-k} = F_{1,n-k}$.

Doble Grado en II+ADE **04/06/ 2019**

Ecometría Examen Final Curso 2018-2019

DNI/NIE : **APELLIDOS:** **Nombre:**

Ejercicio 1(4 puntos). Dados los siguientes datos

y	x_2	x_3
14	3	6
11	2	4
7	1	3
12	4	2
10	3	1
11	2	3

1) Obtener el plano de regresión de y sobre x_2 y x_3.

2) ¿Están las variables x_2 y x_3 fuertemente correlacionadas?

3) ¿Son las variables x_2 y x_3 individualmente significativas?, acotar los p-valores correspondientes.

4) ¿Son las variables x_2 y x_3 globalemente significativas? Suprimiendo variables, ¿cúal es el modelo más significativo?

Ayuda: $\begin{pmatrix} 6 & 15 & 19 \\ 15 & 43 & 46 \\ 19 & 46 & 75 \end{pmatrix}^{-1} = \begin{pmatrix} 2.3298 & -0.5273 & -0.2668 \\ -0.5273 & 0.187 & 0.0189 \\ -0.2668 & 0.0189 & 0.0693 \end{pmatrix}$

1) Con los datos, construimos las matrices

$X = \begin{pmatrix} 1 & 3 & 6 \\ 1 & 2 & 4 \\ 1 & 1 & 3 \\ 1 & 4 & 2 \\ 1 & 3 & 1 \\ 1 & 2 & 3 \end{pmatrix}$, $\mathbf{y} = \begin{pmatrix} 14 \\ 11 \\ 7 \\ 12 \\ 10 \\ 11 \end{pmatrix}$. El método de mínimos cuadrados funcionará ya que el rango de X es 3 (p.ej, tomando el menor de las tres primeras filas).

Entonces:

$X'X = \begin{pmatrix} 6 & 15 & 19 \\ 15 & 43 & 46 \\ 19 & 46 & 75 \end{pmatrix}$, $\hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{pmatrix} = (X'X)^{-1}X'\mathbf{y} = \begin{pmatrix} 6.6399 \\ 1.7849 \\ 0.8587 \end{pmatrix}$,

$y = 6.6399 + 1.7849x_2 + 0.8587x_3$, (plano de regresión).

2) Interpretamos la pregunta como la correlación entre los estimadores de los coeficientes correspondientes. La varianza σ del modelo se estima por

$$s^2 = \frac{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}{n-k} = 0.9641, \text{ya que } \hat{\boldsymbol{\varepsilon}} = \mathbf{y} - X\hat{\boldsymbol{\beta}},\ n-k=3.$$

Por tanto, como la matriz de covarianzas se estima por

$s^2(X'X)^{-1} = s^2(a_{ij}) := (c_{ij})$,

su término $c_{23} = 0.0182$ nos da una estimación de la covarianza de $\hat{\beta}_2$ y $\hat{\beta}_3$. Como este valor es pequeño, no están fuertemente correlacionadas.

3) Tenemos dos tests de significación individuales, uno para cada variable, comparando el modelo completo con el restringido al quitar una de las variables:

a) Test variable x_2: H0: $\beta_2 = 0$, H1: $\beta_2 \neq 0$.

b) Test variable x_3: H0: $\beta_3 = 0$, H1: $\beta_3 \neq 0$.

Calculando los t-valores, a partir de $s_j = \sqrt{c_{jj}}$:

$$s_2 = 0.4246,\ s_3 = 0.2585,\ t_2 = \frac{\hat{\beta}_2}{s_2} = 4.2036,\ t_3 = \frac{\hat{\beta}_3}{s_3} = 3.322.$$

Ambos valores caen en la región $t > 3.1882$ de rechazo del estadístico $t(3)$. Por tanto, ambas variables son significativas al 0.05, pero no demasiado, sobre todo la variable x_3: sus p-valores se encuentran en los siguientes intervalos

t_2: $0.02 < p - valor < 0.03$

t_3: $0.04 < p - valor < 0.05$.

4). El contraste es

H0: $\beta_2 = \beta_3 = 0$

H1: algún β_j es no nulo, $j = 2, 3$

Calculemos el F-valor:

$$R^2 = 0.8922,\ F = \frac{3}{2}\frac{R^2}{1 - R^2} = 12.4156.$$

Este F-valor> 9.552, para $F(2, 3)$, por tanto, las variables son conjuntamente significativas.

Conclusión: el modelo completo es el más significativo, aunque la variable x_3 no es demasiado significativa (céteris-páribus).

Doble Grado en II+ADE **04/06/ 2019**

Ecometría **Examen Parcial** **Curso 2018-2019**

DNI/NIE : **APELLIDOS:** **Nombre:**

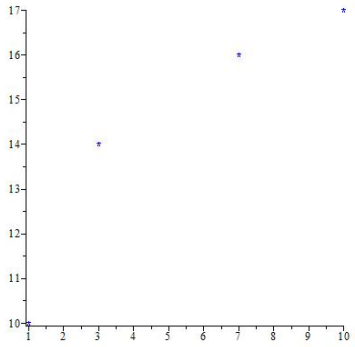
Ejercicio 2 (3 puntos) Una empresa de automóviles desea estudiar la dependencia del número de vehículos vendidos mensualmente, respecto al precio de cada uno de los automóviles. Los datos que se obtienen en un mes de prueba son los siguientes, en donde el precio de los vehículos se mide en unidades de 10^4 euros:

Número	Precio
10	1
14	4
16	7
17	10

1) Mediante mínimos cuadrados obtener una curva que explique el número de vehículos respecto al precio, ajustando razonablemente la nube de puntos (sugerencia: representar previamente la nube de puntos). Dar una estimación del número de vehículos de 120000 euros que se venderían.

2) ¿Es razonable afirmar que el número de vehículos vendidos depende muy significativamente del precio de los mismos? Obtener el resultado tanto mediante un test t como, equivalentemente, mediante un test F.

1) Representando la nube de puntos, con x precio e y el número de vehículos vendidos, queda claro que hay que hacer un cambio de variable logarítmico:



Por tanto, el modelo será de la forma $y = \beta_1 + \beta_2 z + \epsilon$, con $z = \log x$, con lo cual hemos de hacer el cambio de variable $z_j = \log x_j$, con lo que los datos son

$X = \begin{pmatrix} 1 & 0 \\ 1 & 1.3863 \\ 1 & 1.9459 \\ 1 & 2.3026 \end{pmatrix}$, $\mathbf{y} = \begin{pmatrix} 10 \\ 14 \\ 16 \\ 17 \end{pmatrix}$. El método de mínimos cuadrados funcionará ya que el rango de X es 2 (p.ej, tomando el menor de las dos primeras filas).

Entonces:

$X'X = \begin{pmatrix} 4 & 5.6348 \\ 5.6348 & 11.0103 \end{pmatrix}$, $(X'X)^{-1} = \begin{pmatrix} 0.8959 & -0.4585 \\ -0.4585 & 0.3254 \end{pmatrix}$ $\hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = (X'X)^{-1}X'\mathbf{y} = \begin{pmatrix} 9.9445 \\ 3.0564 \end{pmatrix}$, $y = 9.9445 + 3.0564z$, (recta de regresión). Por tanto, mediante mínimos cuadrados obtenemos la curva

$$y = 9.9445 + 3.0564 \log x.$$

El número de vehículos estimados para $x = 12 \cdot 10^3$ euros sería

$$9.9445 + 3.0564 \log 12 = 17.5394.$$

Por tanto, se venderían entre 17 y 18 vehículos, con lo cual no aumentaría mucho las ventas respecto al precio de 10000 euros, si acaso un vehículo más.

2) El test es

H0: $\beta_2 = 0$

H1: $\beta_2 \neq 0$

Mediante test t:

$\hat{\boldsymbol{\varepsilon}} = \mathbf{y} - X\hat{\boldsymbol{\beta}}$, $s = \sqrt{\frac{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}{2}} = 0.155$, $s_2 = s\sqrt{a_{22}} = 0.0884$, siendo $a_{22} = 0.3255$ el término correspondente de la matriz $(X'X)^{-1}$, calculada en 1). Por tanto, el t-valor será

$$t_2 = \frac{\hat{\varepsilon}_2}{s_2} = 34.569.$$

Este es el t-valor de la pendiente. Estamos en t(2), con un t-valor $34.569 >> 4.303$, que es el valor crítico para significación de 0.05, además el p-valor estarápor debajo de 10^{-3}. Por tanto, rechazamos claramente H0: los vehiculos vendidos dependen muy significativamente del precio (o las variaciones en el número de vehículos vendidos se explican muy bien por el precio). Fijémonos que hemos utilizado aquí el término "rechazar H0", en vez de "no podemos aceptar H0", ¿por qué?.

Mediante test F:

$R^2 = 0.9983$, $F = 2\frac{R^2}{1-R^2} = 1195.0199$. Estamos en F(1,2) con un F-valor $1195.0199 >> 18.51$ para 0.05 de significación, mismo resultado que con el test t: rechazamos claramente H0.